

1. Luego de conseguir el archivo de secuencias GenBank de uno de los mRNA de nuestro gen (KIF1B) utilizamos biojava para obtener las traducciones en secuencias de aminoácidos posibles (6 reading frames). Todo eso se guardó en un archivo fasta donde la primera línea de cada secuencia describe el mRNA y el frame del que viene la traducción.
2. Habiendo conseguido todas las traducciones a secuencias de aminoácidos del ejercicio anterior se realizaron 6 consultas BLAST para obtener el resultado de cada frame de lectura. Estos resultados fueron guardados en archivos .out con las consultas del BLAST.

Utilizando los estadísticos como el score, el e-value y el porcentaje de la secuencia que se alineaba pudimos determinar que el frame de lectura correcto era el frame 2. Un e-value menor significa que el alineamiento es más probable a no ser por casualidad. Es la cantidad de hits con score similar que se podrían esperar random, además utilizando el score como medida de mejor alineamiento pudimos encontrar las proteínas similares en distintos organismos. Otra estadística que se puede destacar es la cantidad de gaps que se encontraron al alinear, ya que es más probable que se encuentren mutaciones de un aminoácido por otro a que se encuentren gaps en la evolución.

Los organismos con mejor alineamiento fueron:

- Pan troglodytes - chimpancé (entre otros muchos primates)
 - Zalophus californianus - lobo marino californiano (por debajo de muchos otros primates)
 - Callorhinus ursinus - oso marino ártico
3. Realizando un MSA sobre las secuencias a proteínas pertenecientes a los tres organismos arriba detallados se puede observar que la proteína producida por el chimpancé es idéntica a la proteína encontrada en el homo sapiens. Además se puede observar que las proteínas del lobo marino y del oso marino son idénticas entre sí pero varían en pocos lugares (menos de un 5%) y que esas variaciones son en sí cambios de aminoácidos y no cambios mayores como gaps. Otra cosa a destacar es que en una gran mayoría estos aminoácidos cambiados mantienen las mismas propiedades que los que se encuentran en la proteína producida por el homo sapiens.
 4. Utilizando los reportes de blast obtenidos se pueden recoger las características de los hits y se puede obtener una lista de los que en su descripción o id contengan cierto patrón. El patrón es enviado por parámetro y puede ser vacío si se quieren todos los hits.
 5. Se instaló EMBOSS y se generaron dos scripts, uno para encontrar los ORF de una secuencia de nucleótidos y luego la traducción a secuencias de proteínas posibles. Este procedimiento es muy similar al realizado en el ejercicio 1. Luego se utilizó la base de datos prosite para realizar un análisis de dominios de secuencias de aminoácidos, esto luego se exporta a un archivo donde se encuentra ese análisis de

dominios utilizando la base de datos de motivos de prosite.

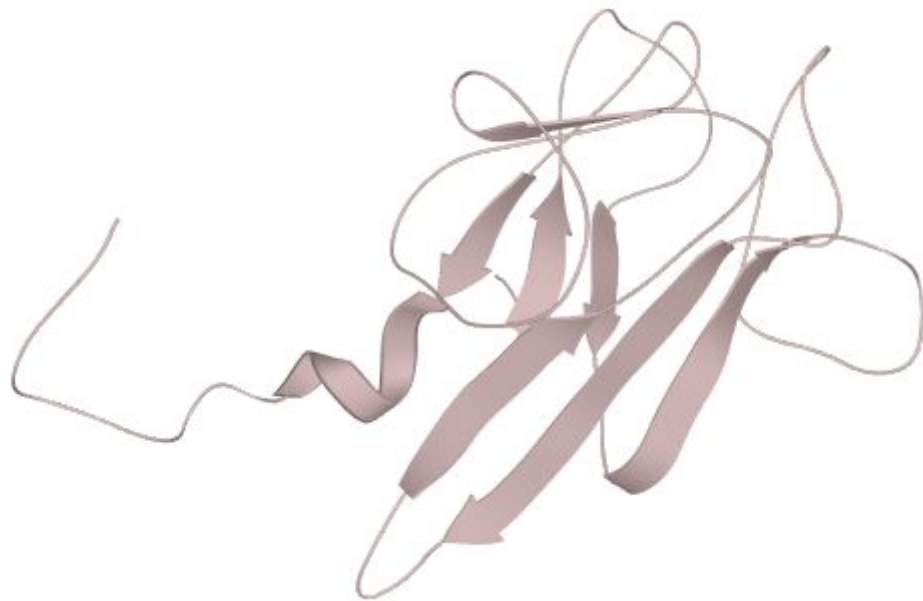
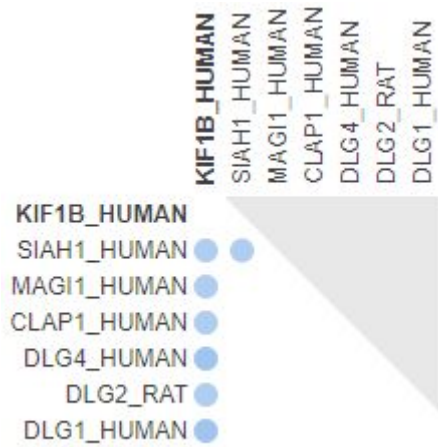
6.

- a. El gen que decidimos elegir es el KIF1B del homo sapiens (<https://www.ncbi.nlm.nih.gov/gene/23095>). Este gen, ubicado en el cromosoma 1 (1p36.22), codifica proteínas motoras que transporta precursores de mitocondrias y vesículas sinápticas además de esto, estas proteínas parecen estar involucradas en el proceso de apoptosis. Por esto resulta muy importante en el funcionamiento correcto de las células. Mutaciones en este gen pueden causar Síndrome de Charcot-Marie-Tooth o incluso susceptibilidad a neuroblastomas. Es por esto que decidimos elegir este gen, nos resultó interesante poder comprender alguna de las causas del neuroblastoma.

Esta enfermedad es el tipo de cáncer de niños más común diagnosticado antes del año de edad y es el responsable de entre 10% y 15% del total de muertes por cáncer en niños. Algunos de los pacientes heredan la predisposición genética mientras que otros lo desarrollan por mutaciones de sus genes.

- b. Dadas las dos bases de datos se puede observar que en la base de datos HomoloGene se encuentran pocos genes homólogos (entre ellos en organismos como roedores, pescados, lobos, primates, gallinas incluso), en cambio en la base de datos de Ensembl se encuentran 199 ortólogos y 41 parálogos. Entre ellos también están los primates, roedores, peces, pájaros y más. Las diferencias entre ambas bases es que en Ensembl se encuentran todos los ortólogos incluso con porcentajes de confianza, en la otra base parecen encontrarse solamente los más cercanos.

Este gen se encuentra en vertebrados y considerando la cantidad de especies que se consiguen como resultados en ambas consultas no parece ser un gen común dentro de los vertebrados.



- e. La proteína KIF1B forma parte de procesos biológicos de transporte intracelular, así como transporte intracelular de vesículas en células neuronales. Participa en movimientos por medio de microtúbulos de vesículas sinápticas y de mitocondrias. Además forma parte del proceso de transmisión sináptica entre neuronas y del proceso de apoptosis. Su función molecular básica es motora, se une a los microtúbulos y a los materiales de transporte de interés (tanto proteínas, como ATP). Los componentes biológicos en los que actúa son por ejemplo el axón, las dendritas, microtúbulos, sinapsis, etc.

- f. Las proteínas producto del gen KIF1B con su mismo nombre participan en dos vías metabólicas importantes. Una es la hemostasia y otra es el transporte mediado por vesículas.

Hemostasia: es la respuesta fisiológica que culmina en la formación de un coágulo para frenar la pérdida de sangre en un vaso sanguíneo. Las proteínas formadas, siendo kinesinas, se unen a los microtúbulos y

transportan materiales en el ciclo de producción de plaquetas en los megacariocitos.

Transporte mediado por vesículas: todo movimiento de proteínas y otros compuestos en las células requiere de vesículas. Las proteínas producto del gen KIF1B transportan estas vesículas, específicamente las relacionadas con mitocondrias y vesículas sinápticas, a través de las células utilizando los microtúbulos como guía.

- g. Existe una variante (rs121908164) asociada a la susceptibilidad al neuroblastoma 1, esta variante es de un solo nucleótido G por A, esto causa un cambio en la traducción de la secuencia de aminoácidos. El cambio es una serina por una asparagina. La frecuencia de la variación en la población es de 0.00001 y afecta mayormente a la población europea.